EEL 6764 0011 **Graduate Computer Architecture** **Sample Exam**
Spring 2023 **Your Name:_____**

Notes:

- Closed Text and Closed Notes Exam
- Time: 3:30 pm – 4:45 pm
- There are **5** questions. Answer all questions.
- Maximum points: 50.
- Answer in clear and legible handwriting. Partial credit will be given.

I	II	III	IV	V	Total

I. (10 pts.) Short questions, Fill in the blanks, True or False.

1. What are power and memory walls in the context of computer architecture?
   Power Wall: As the number of transistors increases, dynamic and static power increases. Excessive power consumption leads to thermal issues and lifetime / reliability reduction. Thus, power consumption limits the performance gains over generations.
   Memory Wall: CPU performance and memory performance improvements are not the same. Therefore, there is a reduction in system performance due to mismatched speeds between CPU and memory.
2. MIPS is an accurate measure of computer performance. True/False
3. Suppose a new processor is 10 times faster in computation than the current one. Assume that the current processor is 40% time busy with computation and 60% of the time idle waiting for IO. What is the overall speedup with the new processor?
   Applying Amdahl's law, speedup = 1 /(0.6 + (0.4/10)) =1.43
4. Briefly state the two principles of locality.
   Spatial Locality – Items nearby an item referenced are likely to be referenced soon.
   Temporal Locality – An item referenced is likely to be referenced soon.
5. A program consisting of 500 instructions is executed on a 5-stage processor. How many cycles would be required to complete the program. Assume ideal overlap in case of pipelining. What is the speedup with pipelining?

Total cycles = Fill latency + 1 instruction per cycle after fill = 5 + 500 = 505
Speed up with the pipelining = No. of stages in the pipeline = 5

6. TLB stands for ____Translation Lookaside Buffer_____

   RISC stands for___Reduced Instruction Set Computer_____

7. RISC-V architecture is what type? _____RISC_____

8. Branches comprise 20% (BF) of all instructions. Branch prediction is 80% accurate (BPA). 2 cycle stalls (SP) on each mis-prediction. Compute average branch penalty control hazards on CPI of the pipelined processor.
   Solution: $Branch\ Penality = BF \times (1 - BPA) \times SP$
   $= 20\% \times (1 - 80\%) \times 2$
   $= 0.2 \times (1 - 0.8) \times 2$
   $= 0.2 \times 0.2 \times 2$
   $= 0.08$
9. In __Critical Word First _____ scheme, we will provide CPU immediately with the instruction that we fetch from lower level of memory (instead of waiting for the entire block to be transferred).

10. Circle appropriately: For non-blocking caches, the bandwidth is ____**better /worse/ no change**_____, miss penalty is __**better / worse**_____, hardware cost ____**low / high / none**_____

II. (10 pts) Multiple choice questions. More than one choice may be correct. Credit only if you choose all correct choices.

a. Which of the following is a true measure of computer performance?
   (a) CPI (c) MIPS
   (b) Clock rate (d) None of the above.

b. Which of the following are power reduction techniques
   (a) Voltage scaling (c) Increasing clock frequency
   (b) Frequency scaling (d) None of the above.

c. The Write Through scheme can be improved by:
   (a) Write Buffer (b) Critical Word first (c) Write-back (d) Dirty blocks

d. Which of the following can improve cache performance?
   (a) Smaller block size (b) Higher Associativity (c) Multi-level Cache (d) None of the above

e. The main advantages of merging write buffer optimization is
   (a) reduces hit-time
   (b) reduces miss penalty
   (c) reduces miss rate
   (d) none of the above.

f. Which of the following page replacement algorithm is popular in processors?
   (a) Least Recently Used (b) Random (c) Most Recently Used (d) None of the Above.

g. To keep page table size under control, we can use

(a) Larger Virtual Address (b) Multi-level Page Table (c) Large physical address (d) None of the above.

h. In an in-order processor, the following hazards are not an issue:
(a) RAW (b) WAW (c) WAR (d) None of the above.

i. The drawbacks of loop unrolling is/are:
(a) register shortfall (b) decrease in code size (c) some ld/sd instruction dependences cannot be identified (d) None of the above.

j. Assuming an in-order pipeline with data forwarding, the following code has:
(a) data hazard (b) control hazard (c) structural hazard (d) None of the above.

```
Loop: fld     f0, 0(x1)
      fadd.d  f4, f0, f2
      fsd     f4, 0(x1)
      addi    x1, x1, -8
      bne     x1, x2, Loop
```

III. Technology Trends, Performance Measurement

a. (4 pts) Describe briefly how the following ISA works?
Stack, Accumulator, Register-memory, register-register. (No need of a figure)

1. **Stack ISA:** A Stack ISA is an instruction set architecture (ISA) that operates on a stack data structure. The instructions in this architecture typically push and pop data onto and from the top of the stack, allowing the processor to manage the flow of data.
2. **Accumulator ISA:** An Accumulator ISA is an ISA that uses a single register, called the accumulator, to hold the result of arithmetic operations. The accumulator is also used as an implicit operand for many instructions.
3. **Register-memory ISA:** A Register-memory ISA is an ISA that uses a combination of registers and memory to store data and perform operations. This type of ISA provides more flexibility than the accumulator ISA, as it allows multiple operands to be used in a single instruction.
4. **Register-register ISA:** A Register-register ISA is an ISA that operates entirely on registers, with no memory access required. All operands and results are stored in registers, providing the fastest possible execution.

b. Calculate the effective CPI for a RISC-V CPU. Assume the following: All ALU Operations – 1 clock cycle, Loads – 5 cycles, stores – 3 cycles, branches taken – 5 branches not taken – 3, jumps – 3
The benchmark data is:

Gcc: 17% loads, 23% stores, 20% branches, 4% jumps, ALU operations – rest.

Solution:
Given ALU Ops = 1 Loads = 5 Stores = 3
Branches taken = 5 Branches not taken = 3 Jumps = 3

Benchmark data:
Loads = 17% Stores = 23% Branches = 20% Jumps = 4% ALU = 36%

Effective CPI = ∑Instruction Clock Cycles * Instruction Frequencies
= 17% * 5 + 23% * 3 + 36% * 1 + 4% * 3 + 20 % ((5 + 3) /2)
= 0.17 * 5 + 0.23 * 3 + 0.36 * 1 + 0.04 * 3 + 0.2 * 4
= 0.85 + 0.69 + 0.36 + 0.12 + 0.8
= 2.82

IV. Memory Hierarchy

(4 pts) Briefly (in 2-3 sentences each) describe **any four** advanced cache optimization techniques. Solution:

a) **Way Prediction:** To predict the way or block inside the set of the upcoming cache access to reduce hit time, extra bits are retained in the cache. The multiplexor is set early to choose the appropriate block as a result of this prediction, and just one tag comparison is carried out in parallel with reading the cache data. In the event of a miss, the following clock cycle will search the other blocks for matches.

b) **Nonblocking Cache:** For pipelined computers that support out-of-order execution, the processor must halt on a data cache miss. By enabling the data cache to keep supplying cache hits even in the event of a miss, a nonblocking cache or lockup-free cache increases the potential benefits and increases cache bandwidth. By assisting during a miss rather than disregarding the processor's demands, this "hit under miss" optimization lowers the effective miss penalty.

c) **Critical Word First:** This method is based on the observation that the processor typically only requires one word at a time from the block. To fill in the remaining words in the block, ask for the missing word first from memory and transmit it to the processor as soon as it is available. Then, let the processor carry on with its current task.

d) **Merging Write Buffer**: Because every store needs to be transmitted to the next lower level of the hierarchy, write-through caches rely on write buffers to reduce miss penalty. The addresses of the new data can be compared to the addresses of valid write buffer entries as the buffer is filled. If so, fresh information is added to that record.

(6 pts) A Cache acts as a filter. For example, for every 1000 instructions of a program, an average of 20 memory accesses may exhibit low enough locality that they cannot be services by a 2 MB cache. The 2MB cache is said to have an MPKI (misses per thousand instructions) of 20, and this will be largely true regardless of the smaller caches that precede the 2MB cache. Assume the following cache/latency/MPKI values: 32KB/1/100, 128/2/80, 512KB/4/50, 2MB/8/40, 8MB/16/10. Assume that accessing the off-chip memory system requires 200 cycles on average. For the following cache configuration, calculate the average time spent accessing the cache hierarchy: 32KB L1; 8 MB L2; off-chip memory.

Solution : AMAT = (L1 misses) * 16 cycles + (L2 misses) * 200

= 100 * 16 + 10 * 200
= 1600 + 2000
= 3600 cycles

V. Basic Pipelining and RISC-V architecture

(6 pts) A. Suppose the branch frequencies (as percentages of all instructions) is as follows: Conditional Branches: 15% Jumps/Calls: 1% Taken Conditional Branches: 60% Taken We are examining a 4-stage pipeline where the branch is resolved at the end of the second cycle for unconditional branches and at the end of the third cycle for the conditional branches. Assuming that only the first pipe stage can always be completed independent of whether the branch is taken and ignoring other pipeline stalls, how much faster would the machine be without any branch hazards.

Solution:

The performance of ideal pipeline without branch hazards is the pipeline depth, which is 4. Next, for branch hazards, we need to figure out the stall cycles for each type of branches.

- For unconditional branches, the stall cycle is 1.
- For conditional branches which is taken, the stall cycles is 2.
- For conditional branches which is not taken, the stall cycle is 1.

Considering the frequencies of different types of branches,

The performance of pipeline with branch hazards = (Pipeline depth)/(1+pipeline stalls)
= 4/(1+(1 ×1% + 2 × 15% × 60% + 1 × 15% × 40%) = 3.2

Speedup = (Pipeline Speedup without Hazards)/(Pipeline Speedup with Hazards)
= 4/3.2=1.25

(4 pts) B. Answer the following RISC-V Base ISA architecture:

a) Name any *four ISA* design principles that RISC-V implements.

b) No. of integer registers _____ No. of FP registers_____

c) Instruction word length_____

d) No. of instruction formats _____

Solution:

a) ISA design principles
   - Simplicity: It is easy to design and implement and simple to understand.
   - Modularity: It is designed in a modular way so that when the designer needs any specific set of instructions, it can use the extension that is provided by the RISC.
   - Orthogonally: It is very flexible as the small set of instructions can be combined to form the complex instructions so that it will be able to execute either ways.
   - Extensibility: Whenever designer needs to add any feature that can add very easily without disturbing the existing functionality.

b) 32, 32

c) 32-bit instruction word length

d) 4. I – Immediate Instruction, R – Register Instruction, S – Store Instruction, U – Jump Instruction